

Эволюция многоядерных гетерогенных вычислительных систем в области обработки видеоданных

А.А. Беляев, В.С. Гаврилов, Д.А. Кузнецов, Я.Я. Петричкович, Т.В. Солохина,
Д.С. Фролов, А.А. Функнер

ОАО НПЦ «ЭЛВИС», bel@elvees.com

Аннотация — Рассматриваются основные направления и перспективы развития многоядерных гетерогенных вычислительных систем на кристалле (СнК) в области цифровой обработки сигналов – в первую очередь в области обработки видеосигнала, требующей чрезвычайно высокой скорости вычислений. Дается сравнительный анализ возможностей и перспектив применения различных вычислительных архитектур в рассматриваемой прикладной области.

Ключевые слова — многоядерные гетерогенные системы на кристалле (СнК), цифровая обработка сигналов, компьютерное зрение, видеоаналитика, проблемно-ориентированные процессоры.

I. ВВЕДЕНИЕ

В настоящее время наблюдается очередная смена парадигмы в проектировании сложных СнК, происходящая вследствие, с одной стороны, возрастающих требований к производительности систем обработки данных, с другой стороны, не прекращающегося роста числа транзисторов, которые могут быть размещены на одном кристалле [1, 2].

У нынешнего перехода есть свои отличия. Впервые успехи полупроводникового производства не приводят к пропорциональному росту производительности. Становится очевидно, что, начиная с технологии 65 нм и ниже, только малая часть роста производительности вычислительных систем на кристалле будет связана непосредственно с уменьшением геометрических размеров элементов, большая же часть должна достигаться благодаря архитектурным инновациям.

В связи с этим последние годы ознаменовались фундаментальными архитектурными сдвигами – в первую очередь переходом от одноядерных архитектур к *многоядерным*.

Однако сам по себе переход от последовательного вычислительного процесса к умеренно параллельному ещё не гарантирует того роста производительности, который необходим в наиболее вычислительно-интенсивных прикладных областях. Поэтому другой важной тенденцией последнего времени является *углубление специализации* вычислительных устройств и переход к *гетерогенным* вычислительным системам.

При этом, вследствие того, что методы и алгоритмы обработки данных постоянно совершенствуются, важнейшим требованием остается сохранение *программируемости* таких устройств.

Столь глубокие сдвиги привели к новому витку эволюционной гонки с участием различных типов и вариантов процессорных архитектур.

Наиболее ярко и динамично эта гонка разворачивается в прикладной области, связанной с видеопереработкой. Эта «экологическая ниша» обладает особой значимостью и привлекательностью, поскольку преобладающая часть потребляемой человечеством информации имеет визуальный характер, вследствие чего на хранение, передачу и обработку этой информации затрачиваются огромные ресурсы. Однако сложность и огромный объем вычислений, связанные с обработкой видеоданных, столь велики, что далеко не для всякой архитектуры они оказываются «по зубам».

В данной статье рассматриваются современные тенденции и перспективы применения различных процессорных архитектур в области видеопереработки.

II. ОСОБЕННОСТИ ПРИКЛАДНОЙ ОБЛАСТИ ОБРАБОТКИ ВИДЕОДАНЫХ

Обработка видеоданных, по сравнению с другими видами сигнальной обработки, обладает рядом существенных особенностей, которые должны учитываться при выборе аппаратно-программных платформ для ее реализации:

– основная трудность реализации приложений видеопереработки связана с огромным потоком поступающих данных и сложностью применяемых алгоритмов обработки. Например, для видеосигнала в формате Full HD (1080p 60 fps) поток данных составляет около 120 Mpixel/s, а сложность алгоритмов может составлять 2-3 тысячи операций на пиксель и более – следовательно, вычислительная платформа для видеоприложений должна обладать производительностью, как минимум, в *200-300 GOPs*;

– алгоритмы обработки видеоданных сложны и быстро развиваются – следовательно, важнейшим

требованием к вычислительной платформе для их реализации является *программируемость*;

– алгоритмы обработки видеоданных очень многообразны. В отличие от области традиционной сигнальной обработки, в которой преобладающее значение имеют такие базовые процедуры, как линейная фильтрация и быстрое преобразование Фурье, арсенал алгоритмов видеоаналитики гораздо более широк и разнообразен. Некоторые из них требуют обработки данных в форматах не только с *фиксированной*, но и с *плавающей точкой*.

III. ЭВОЛЮЦИЯ ПРОЦЕССОРОВ ОБЩЕГО НАЗНАЧЕНИЯ В НАПРАВЛЕНИИ ПОДДЕРЖКИ МУЛЬТИМЕДИЙНЫХ ПРИЛОЖЕНИЙ

Процессоры общего назначения (GPP – General Purpose Processor) в силу своей универсальности могут быть использованы для решения любой вычислительной задачи. Однако и среди них имеется архитектурная дифференциация, связанная с типом устройства, для которого предназначен процессор. Процессоры для стационарных компьютеров – это, как правило, процессоры с CISC-архитектурой, а процессоры для встраиваемых применений (embedded processors) – это обычно процессоры с RISC-архитектурой.

Ведущими мировыми производителями процессоров для стационарных компьютеров являются фирмы Intel [3] и AMD [4]; основные производители процессоров для встраиваемых применений – фирмы ARM [5] и Imagination [6].

С целью повышения производительности во многих процессорах общего назначения применяются расширения системы инструкций, в частности:

Intel – SSE (Streaming SIMD Extension), SSE2, SSE3 [7];

ARM – Neon [8];

MIPS – DSP ASE [9].

Таким образом, даже среди процессоров общего назначения явно прослеживается тренд в направлении специализации и поддержки вычислительно интенсивных приложений мультимедийной обработки.

Другим способом *специализации* и повышения производительности процессоров общего назначения является *интеграция с аппаратными ускорителями*. Главный недостаток решений с использованием акселераторов, по сути, тот же, что и вообще для специализированных схем – недостаточная гибкость.

Применение процессора в сочетании со встроенной программируемой логикой (ПЛИС) позволяет частично снять указанный недостаток. Оба устройства – процессор и ПЛИС – позволяют, в принципе, проводить оптимизацию системы даже после изготовления кристалла. Однако решения на основе ПЛИС уступают по производительности специализированным схемам, и, к тому же, довольно дорогостоящи.

IV. ГРАФИЧЕСКИЕ ПРОЦЕССОРЫ В ЗАДАЧАХ ВИДЕООБРАБОТКИ

В тех прикладных областях, для которых производительность процессоров общего назначения оказывается явно недостаточной, применяются специализированные процессоры, ориентированные на реализацию некоторого заданного круга прикладных задач. Такие процессоры называются проблемно-ориентированными (ASIP – Application-Specific Instruction-set Processor). Проблемно-ориентированные процессоры являются высокопроизводительной и энергетически эффективной альтернативой процессорам общего назначения. Они сочетают в себе гибкость программируемых устройств и эффективность специализированных схем.

Одной из наиболее распространенных разновидностей проблемно-ориентированных процессоров являются графические процессоры (GPU – Graphic Processing Unit). Они выделились в особый класс процессорных устройств уже в 90-х годах прошлого века, когда стало очевидно, что решаемые ими задачи формирования изображения на компьютерном терминале требуют сверхвысокой производительности, и, как следствие, применения особой высокопараллельной архитектуры.

На сегодняшний день графические процессоры обладают наивысшей производительностью среди программируемых устройств. По этой причине вполне оправданными представляются попытки применения графических процессоров для решения задач видеоаналитики и обработки видеоданных. Применение графических процессоров для этих задач выглядит тем более оправданным, что обе прикладные области – графика и видеоаналитика – имеют дело с визуальным изображением, хотя решаемые ими задачи до известной степени обратны друг другу: в графических приложениях содержание сцены известно заранее, требуется максимально правдоподобно сформировать соответствующее изображение; в приложениях видеоаналитики, напротив, исходным является само изображение и требуется определить его содержание.

Активно продвигает свои графические процессоры на рынок машинного зрения компания Nvidia. Её последней разработкой в этой области является чип Tegra K1 [10], представляющий собой систему на кристалле, в которую входят массив потоковых процессоров, аналогичных используемых в видеокартах, и кластер ядер ARM. Пиковая производительность системы ~ 326 GFLOPS.

С помощью графических процессоров удалось реализовать некоторые видеоприложения, в частности, стандартные видеокодеки – JPEG, MPEG2, MPEG4, H.264, VP6, VP8 и другие. Однако в целом задачи видеоаналитики, по сравнению с графическими приложениями, характеризуются гораздо большей нерегулярностью, и использование для их решения графических процессоров зачастую оказывается энергетически неэффективным.

V. УГЛУБЛЕНИЕ СПЕЦИАЛИЗАЦИИ СИГНАЛЬНЫХ ПРОЦЕССОРОВ. ПРОБЛЕМНО-ОРИЕНТИРОВАННЫЕ ПЛАТФОРМЫ ДЛЯ ПРИЛОЖЕНИЙ ВИДЕООБРАБОТКИ

Другой широко распространенной разновидностью проблемно-ориентированных процессоров являются цифровые сигнальные процессоры (DSP - Digital Signal Processors). Однако обработка сигналов – весьма широкая прикладная область. В зависимости от типа обрабатываемого сигнала требования к процессору могут сильно отличаться, по этой причине в настоящее время отчетливо проявляется тенденция к углублению специализации процессоров сигнальной обработки.

Это выражается, прежде всего, в быстром развитии процессорных платформ, специализирующихся на обработке видеоданных. Ниже приведен краткий обзор программируемых архитектур для видеообработки, представленных основными фирмами-разработчиками в данной области в последние годы.

Texas Instruments. Фирма Texas Instruments для процессоров семейства DaVinci DM6446, DM6441, DM647, DM648 разработала в 2009 году сопроцессор VICP (Video Image Coprocessor) с архитектурой SIMD и производительностью до восьми 16-разрядных операций умножения с накоплением (MAC-операций) за такт. Для этого сопроцессора была создана библиотека Signal Processing Library для поддержки функций компьютерного зрения OpenCV.

В 2012 году для решения задачи Back over Prevention (BoP) – одной из задач системы автоматического помощника водителя ADAS (Automatic Driver Assistance System) – фирма Texas Instruments разработала другой специализированный процессор – EVE (Embedded Vision Engine). В его состав входит (рис. 1) векторный сопроцессор VCOP (Vector Coprocessor), в котором применяется более широкое 512-разрядное SIMD-распараллеливание. Впервые в этой разработке Texas применил свой собственный RISC-процессор ARP32 (32-bit Application-Specific RISC Processor), поскольку в данной задаче требуется RISC-процессор, более тесно связанный с векторным сопроцессором, чем это возможно при использовании процессоров ARM [11].

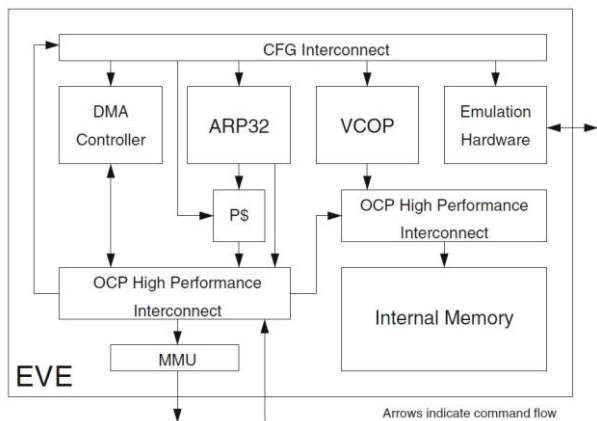


Рис. 1. Архитектура процессора EVE (Embedded Vision Engine) фирмы Texas Instruments

Qualcomm. Фирма Qualcomm для своих процессоров семейства Snapdragon разработала 64-разрядное DSP-ядро Hexagon [12]. Проведенная при его проектировании оптимизация позволила на 10-40% улучшить основные характеристики (потребление и быстродействие) по сравнению с комбинацией ARM + Neon. Для этого процессора была разработана библиотека функций компьютерного зрения FastCV.

Freescale & CogniVue. В процессорах SCP220x (Image Cognition Processors) фирма Freescale использовала IP-ядра APEX, разработанные фирмой CogniVue (рис. 2). APEX – массивно-параллельная архитектура с 96-ю 16-разрядными каналами обработки [13].

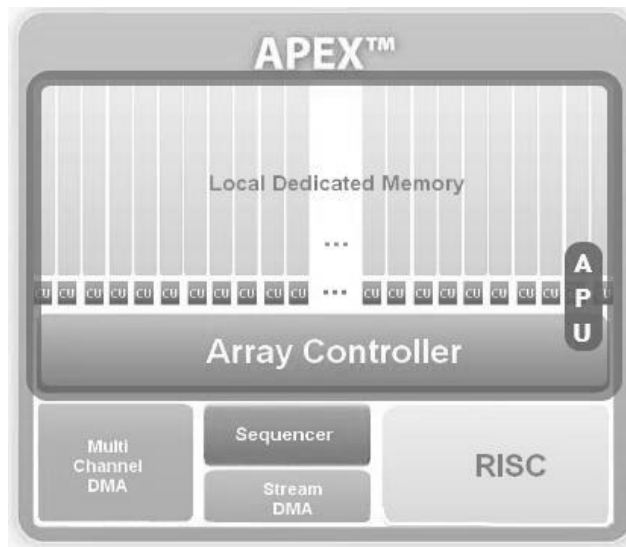


Рис. 2. Архитектура IP-ядра APEX фирмы CogniVue

Cadence & Tensilica. Другая массивно-параллельная архитектура со 128-ю каналами обработки (рис. 3) была предложена фирмой Tensilica (филиал фирмы Cadence) в виде процессорного ядра IVP (Imaging/Video Processor) [14].

При впечатляющих уровнях параллелизма слабость массивно-параллельных архитектур заключается в том, что при реализации нерегулярных, не поддающихся распараллеливанию алгоритмов (а в задачах видеоаналитики такого рода процедур используется довольно много) они оказываются крайне неэффективными. По этой причине при построении многопроцессорных систем видеообработки предполагается, что массивно-параллельные вычислители производят массовую первичную обработку данных, а заключительную нерегулярную обработку выполняют RISC-процессоры.

Проблема заключается в том, что соотношение объемов первичной и вторичной обработки может быть различным и заранее неизвестным, и в случаях преобладания вторичной обработки эффективность массивно-параллельных архитектур снижается.

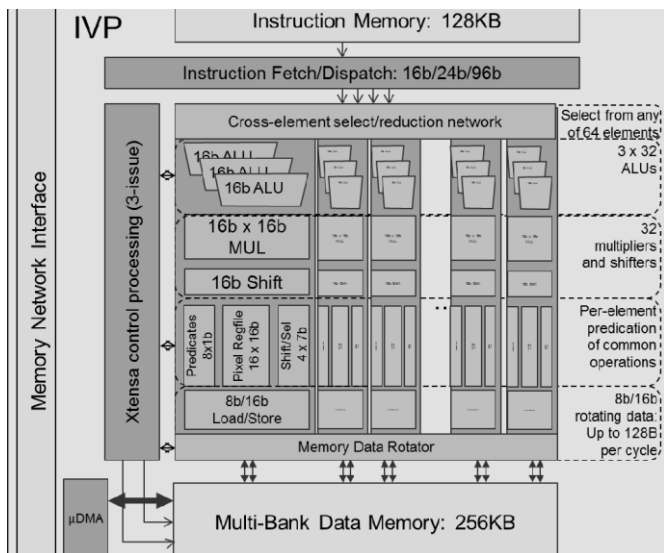


Рис. 3. Архитектура IP-ядра IVP (Imaging/Video Processor) фирмы Tensilica

CEVA и Videantis. Это обстоятельство, по-видимому, учитывают в своих разработках фирмы CEVA [15] и Videantis [16], предлагающие довольно похожие друг на друга решения. Каждая из указанных платформ включает в себя RISC-подобный потоковый процессор (Stream Processor) и векторный DSP-процессор с умеренной степенью SIMD-распараллеливания. Из сочетаний двух этих типов процессоров формируются многоядерные гетерогенные конструкции, причем количество RISC-ядер может быть довольно значительным – в системе CEVA MM-3000, например, используется целых шесть RISC-ядер – два в векторном процессоре и четыре – в потоковом (рис. 4).

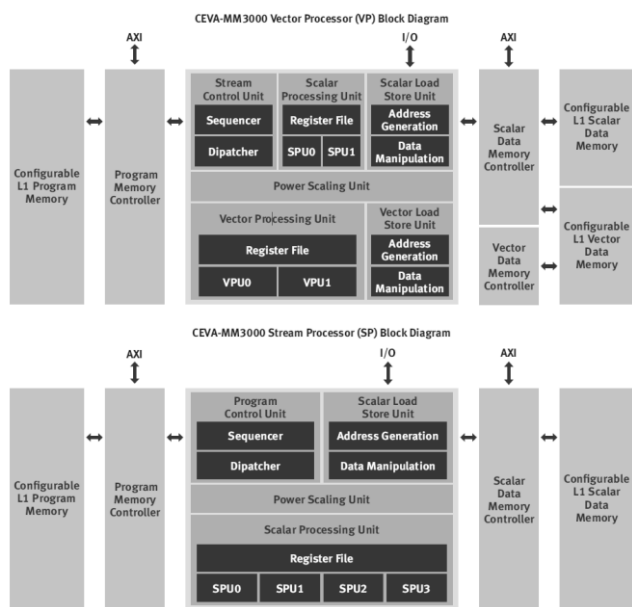


Рис. 4. Архитектура IP-ядер VP (Vector Processor) и SP (Stream Processor) фирмы CEVA

В табл. 1 приведены основные характеристики рассмотренных выше программируемых платформ для обработки видеоданных, а также характеристики рассматриваемой в следующем разделе платформы «Мультикор».

Таблица 1

Сравнительные характеристики программируемых платформ для приложений видеобработки

Компания	Платформа	Количество типов ядер процессоров	Производительность, оп/такт
ARM	ARM + Neon	1	16
Texas Instruments	DaVinci Video Image Coprocessor	1 (VICP)	16
Texas Instruments	Embedded Vision Engine (EVE)	2 (VCOP + ARP32)	32
Qualcomm	Hexagon DSP	1	16
Freescale + CogniVue	APEX	2 (APEX + RISC)	96
Cadence + Tensilica	Imaging/Video Processor (IVP)	2 (IVP + RISC)	128
CEVA	MM 3000	2 (2*VPU + 6*SPU)	32
Videantis	v-MP4280HDX	2 (8*v-MP + 2*v-SP)	192
Elvees	VELCore-02	1 (8*DSP Elcore-40)	384

VI. РАЗВИТИЕ ПЛАТФОРМЫ «МУЛЬТИКОР» В НАПРАВЛЕНИИ ПОДДЕРЖКИ ПРИЛОЖЕНИЙ ВИДЕОАНАЛИТИКИ

Проводимое в НПЦ «ЭЛВИС» развитие и углубление специализации аппаратно-программной платформы «Мультикор» [17] привело к созданию специализированной платформы VIPP (Video Intelligent Processing Platform) для поддержки приложений видеоналики.

Архитектура созданной в рамках VIPP-платформы многоядерной гетерогенной вычислительной системы VELCore-02 для задач видеобработки представлена на рис. 5.

Предлагаемая архитектура включает в себя:

- процессорные ядра сигнальной (видео) обработки DSP0-DSP7 (каждое ядро имеет 2 шины – X и Y – для доступа к памяти данных и одну шину – P – для доступа к памяти программ);

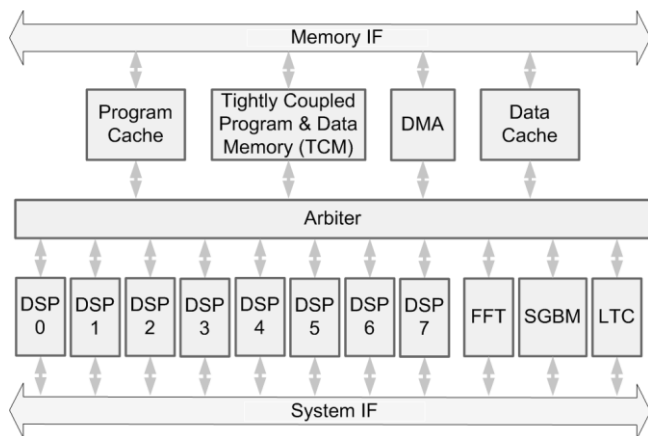


Рис. 5. Архитектура многоядерной гетерогенной вычислительной системы VELCore-02 для приложений видеобработки

- общую (доступную для всех процессорных ядер и аппаратных ускорителей) память программ и данных TCM;

- многоканальный DMA-контроллер для массовых перемещений данных между памятью, входящей в состав системы, и внешней памятью;

- кэш программ (для программ, хранящихся во внешней памяти) Program Cache;

- кэш данных (для данных, хранящихся во внешней памяти) Data Cache;

- аппаратные ускорители различного назначения для реализации критически важных алгоритмов видеобработки (каждый из них имеет доступ к памяти данных), в том числе: FFT (Fast Fourier Transform) – ускоритель быстрого преобразования Фурье; SGBM (Semi-Global Block Matching) – ускоритель алгоритма построения карты глубины для стерео зрения; LTC (Lossy/Lossless Tile Compressor) – блок компрессии/декомпрессии видеоданных;

- стандартные шины AMBA AHB и AMBA AXI для встраивания предлагаемой вычислительной системы в систему-на-кристалле.

Предлагаемая платформа обладает следующими базовыми свойствами:

- *гетерогенность*. Использование в рамках одной архитектуры процессорных ядер различного типа (от RISC до специализированных процессорных ядер), а также аппаратных ускорителей, позволяет использовать преимущества тех и других для достижения эффективности и гибкости вычислений;

- *многопоточность*. Все входящие в систему вычислительные ядра могут функционировать одновременно, обеспечивая высокую скорость вычислений;

- *общая* (для всех ядер) *статическая память* программ и данных позволяет гибко перераспределять ресурсы памяти между ядрами и решает проблему

обмена данными между ними. Сами процессорные ядра при этом сильно упрощаются – в их составе остаются, в основном, лишь регистровый файл, АЛУ и устройство программного управления. Площадь таких ядер невелика, их можно разместить на кристалле в довольно большом количестве. Сложность такого подхода состоит в топологической реализации общей памяти и обеспечении требуемой скорости обмена;

- *общие кэши* программ и данных. Применение общего для всех ядер кэша данных (в отличие от использования индивидуальных кэшей для каждого ядра) решает проблему когерентности и позволяет гибко и динамично перераспределять его ресурсы. Кроме того, в этом случае отпадает необходимость в кэше 2-го уровня. Потенциальный недостаток состоит в перегруженности кэша в случае одновременных обращений со стороны нескольких ядер. Во избежание этого наиболее массовые перемещения данных между внешней и внутренней памяти должны производиться с использованием механизма DMA, кэш-память в этом случае рассматривается как дополнительный механизм, обеспечивающий гибкий доступ к полному адресному пространству при одиночных обращениях. Таким образом, механизмы кэш-памяти, с одной стороны, и статической памяти плюс DMA, с другой, взаимно дополняют друг друга, что позволяет гибко перераспределять между ними трафик обращений к внешней памяти;

- *открытость*. Предлагаемая архитектура открыта для будущих усовершенствований применительно к различным прикладным областям и использования в ее составе процессорных ядер и ускорителей различного назначения. Единственным требованием к используемым устройствам является соблюдение правил доступа к общей памяти. Принцип открытости и наращиваемости распространяется также и на систему инструкций процессорных ядер, разрабатываемых для VIPP-платформы – в частности, в систему инструкций могут быть включены макрокоманды для управления аппаратными ускорителями, запускающие сложные механизмы обработки данных;

- *масштабируемость*. Могут быть изменены параметры как отдельных ядер, так и системы в целом, в том числе: количество процессорных ядер и ускорителей, объемы памяти и разрядность шин данных и др.;

- *использование стандартных интерфейсов*. Этот общий принцип построения многоядерных систем на кристалле в полной мере соблюдается в предлагаемой архитектуре. Для встраивания в систему на кристалле используются шины AMBA APB и AMBA AXI.

Архитектура многоядерной гетерогенной системы VELCore-02 предоставляет пользователю следующие преимущества:

- программируемая платформа с поддержкой стандарта OpenVX;

- аппаратные ускорители для критических вычислений позволяют экономить время и мощность;
- использование одного типа программируемого ядра позволяет избежать проблем переноса кода между ядрами;
- программируемое ядро поддерживает все типы данных языка C, включая типы данных с плавающей запятой, чем обеспечивается эффективная работа C компилятора;
- совместная работа программируемых ядер и аппаратных ускорителей на общем поле памяти позволяет достичь чрезвычайно высокого отношения производительность/площадь.

В НПЦ «ЭЛВИС» было выполнено схемотехническое проектирование и проведен синтез многоядерной гетерогенной системы VELCore-02 по технологии TSMC 40 nm LP. В табл. 2 приведены полученные в результате физические характеристики и характеристики производительности ядра VELCore-02 в сравнении с IP-ядрами v-MP4280HDX (Videantis) и Cortex-A9 Dual (ARM).

Таблица 2

Сравнительные характеристики IP-ядер Velcore-02, v-MP4280HDX, Cortex-A9 Dual

	Company		
	ELVEES	Videantis	ARM
IP-ядро	Velcore-02	v-MP4280HDX	Cortex-A9 Dual
Частота, МГц	750	400	1000
Площадь, мм ²	12.4	4.28	6.2
Производительность, GFLOPs/GOPs:			
32-р. плав.	51.2	-	4.0
32-р. фикс.	51.2	12.8	4.0
16-р. фикс.	204.8	51.2	8.0
8-р. фикс.	307.2	76.8	16.0
Производительность на мм ² :			
32-р. плав.	4.13	-	0.64
32-р. фикс.	4.13	3.0	0.64
16-р. фикс.	16.52	12.0	1.28
8-р. фикс.	24.78	17.9	2.56

VII. ВЫВОДЫ

Проведенный анализ современных тенденций развития многопроцессорных гетерогенных архитектур в области обработки видеоданных позволяет сделать следующие выводы:

- быстрая эволюция процессорных архитектур в области обработки видеоданных, происходящая в

настоящее время, определяется ее особой сложностью и значимостью;

- в настоящее время преобладающее положение в данной области занимают программируемые платформы, основанные на применении двух типов процессорных ядер – массивно-параллельного ядра для высокопроизводительных регулярных вычислений и RISC-подобного ядра для нерегулярных процедур;

- в НПЦ «ЭЛВИС» в рамках развития платформы «Мультикор» создана специализированная платформа VIPP (Video Intelligent Processing Platform) для поддержки приложений по обработке видеоданных, основанная на применении однотипных универсальных DSP-ядер с поддержкой функций стандарта OpenVX и аппаратных ускорителей, работающих на общем поле памяти. и обладающая существенными преимуществами: чрезвычайно высокой достигаемой производительностью в расчете на единицу площади кристалла; эффективной поддержкой компилятора языка C; простотой переноса программного кода между ядрами вследствие использования только одного типа программируемого ядра.

ЛИТЕРАТУРА

- [1] Kempf T., Ascheid G., Leupers R. Multiprocessor Systems-on-Chips. Design Space Exploration. Springer, 2011.
- [2] Hübner M., Becker J. Multiprocessor System-on-Chip: Hardware Design and Tool Integration. Springer, 2010.
- [3] <http://www.intel.com> (дата обращения: 24.03.2014).
- [4] <http://www.amd.com> (дата обращения: 24.03.2014).
- [5] <http://www.arm.com> (дата обращения: 24.03.2014).
- [6] <http://www.mips.com> (дата обращения: 24.03.2014).
- [7] Intel®64 and IA-32 Architectures Software Developer's Manual. Intel Corporation, 2011.
- [8] <http://www.arm.com/products/processors/technologies/neon.php> (дата обращения: 24.03.2014).
- [9] MIPS32® Architecture For Programmers. Volume IV-e: The MIPS® DSP Application-Specific Extension to the MIPS64® Architecture. MIPS Technologies. April 06, 2010.
- [10] <https://developer.nvidia.com/jetson-tk1> (дата обращения: 28.03.2014).
- [11] <http://www.ti.com/lit/wp/spry251/spry251.pdf> (дата обращения: 26.03.2014).
- [12] <https://rtpw.io/slides/rpw-pacsec2013-hexagon.pdf> (дата обращения: 26.03.2014).
- [13] <http://www.cognivue.com/technology.php> (дата обращения: 26.03.2014).
- [14] <http://ip.cadence.com/ipportfolio/tensilica-ip/image-video-processing> (дата обращения: 28.03.2014).
- [15] CEVA-MM3000. Product Brief. CEVA, Inc., 2011.
- [16] <http://www.videantis.com/products/hardware-ip-cores/unified-platforms> (дата обращения: 28.03.2014).
- [17] Петричкович Я.Я., Солохина Т.В. Цифровые сигнальные контроллеры «МУЛЬТИКОР» - новые отечественные серии систем на кристалле // Доклады 6-й Международной конференции «Цифровая обработка сигналов и ее применения (DSPA-2004)». Москва, 31 марта - 2 апреля 2004. Т. 1. С. 8-15.